

# An Effective Marketing Strategy for Revenue Maximization with a Quantity Constraint

Ya-Wen Teng <sup>#1</sup>, Chih-Hua Tai <sup>\*2</sup>, Philip S. Yu <sup>†3</sup>, Ming-Syan Chen <sup>#4</sup>

<sup>#</sup> *EE Dept., National Taiwan University, Taipei, Taiwan*

<sup>1</sup> [ywteng@arbor.ee.ntu.edu.tw](mailto:ywteng@arbor.ee.ntu.edu.tw), <sup>4</sup> [mschen@cc.ee.ntu.edu.tw](mailto:mschen@cc.ee.ntu.edu.tw)

<sup>\*</sup> *CSIE Dept., National Taipei University, New Taipei, Taiwan*

<sup>2</sup> [hanatai@mail.ntpu.edu.tw](mailto:hanatai@mail.ntpu.edu.tw)

<sup>†</sup> *CS Dept., University of Illinois at Chicago, Chicago, IL 60607*

<sup>3</sup> [psyu@uic.edu](mailto:psyu@uic.edu)

## ABSTRACT

Recently the influence maximization problem has received much attention for its applications on viral marketing and product promotions. However, such influence maximization problems have not taken into account the monetary effect on the purchasing decision of individuals. To fulfill this gap, in this paper, we aim for maximizing the revenue by considering the quantity constraint on the promoted commodity. For this problem, we not only identify a proper small group of individuals as seeds for promotion but also determine the pricing of the commodity. To tackle the revenue maximization problem, we first introduce a strategic searching algorithm, referred to as Algorithm PRUB, which is able to derive the optimal solutions. After that, we further modify PRUB to propose a heuristic, Algorithm PRUB+IF, for obtaining feasible solutions more efficiently on larger instances. Experiments on real social networks with different valuation distributions demonstrate the effectiveness of PRUB and PRUB+IF.

## 1. INTRODUCTION

Due to the advance of the Web 2.0 techniques, various kinds of websites have become parts of our life. Most people nowadays are used to sharing, seeking and obtaining information, and interacting with others through various social networking websites such as Digg, Slashdot, Facebook, and Twitter, to name a few. As more and more people join these social websites, the phenomenon of immense quantity of information flow and influence spread becomes prominent. This then motivates the utilization of this phenomenon, known as the “word-of-mouth” effect, to bring significant potential benefits in many types of business such as viral marketing and innovation promotion. For example, a company can effectively promote its brand and new products by involving a group of influential people spreading the good words over social websites. Hence, in very recent years, there is active research [3, 7, 15, 19] exploring the problem of influence maximization.

Existing works studied the influence maximization problem generally based on two kinds of diffusion models, the Linear Threshold (LT) and the Independent Cascade (IC) models. In the LT model, the influence strength between a pair of individuals is modeled as a fixed weight, and the influence on an individual is described as cumulative. An individual is activated when the sum of the weights propagated from his/her activated neighbors exceeds his/her own threshold. The IC model assumes that each influence of one to another is an independent chance to activate the latter with a probability. On both models, the problem of influence maximization aims at identifying a proper group (of a specified size) of individuals as seeds to have the number of activated people maximized. However, note that such influence maximization problems have not taken into account the monetary aspect, which usually plays a crucial factor in people’s purchase decisions. That is, maximizing the influence spread does not necessarily bring maximum revenue, as an individual may be very interested in a new smartphone, but does not purchase the smartphone due to its high price.

According to the studies in the fields of social psychology [2, 8] and economics [10, 14, 20], one’s valuation towards a commodity will be positively influenced by those that have purchased the commodity, which is known as the herd mentality or positive network externalities. In other words, people tend to imitate others in their consumer behaviors, which will result in irrational valuations of the commodity, i.e., higher valuations beyond their inherent valuations. Based on these phenomena, Mirrokni et al. [17] addressed the **Revenue Maximization** problem by incorporating the monetary concept into the maximization problem of influence spread based on the LT model. In the monetary LT model, the influence strength (or said weight) is transformed into the valuation concept, and the individual threshold is regarded as the inherent valuation. One’s valuation of a commodity is thus the sum of his/her inherent valuation and the valuation increment from social influences. An individual then pays for a commodity and propagates his/her influence whenever her/his valuation is larger than or equal to the pricing of the commodity. Consider the network shown in Figure 1 as an example, where the number contained in each node is the inherent valuation and the number on an edge from one node to another is the weight, i.e., the influence strength. For the ease of illustration, without loss of generality, the function  $F(x) = x$  is used to transform the effects of the social influences cumulated on an individual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

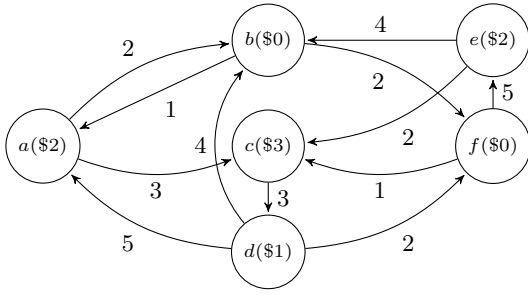


Figure 1: A network with the monetary concept.

into the valuation concept in this example. If  $f$  purchases the commodity,  $c$ 's valuation is increased to  $\$3 + F(1) = \$4$ . If  $a$ ,  $e$ , and  $f$  all purchase the commodity, the valuation increment for  $c$  is  $F(3 + 2 + 1) = \$6$  and  $c$ 's valuation thus becomes  $\$3 + \$6 = \$9$ . The objective of the revenue maximization problem thus aims at determining the pricing of the commodity and identifying a seed group of individuals for promotion so that the total revenue is maximized.

However, to the best of our knowledge, none of the existing works has taken a quantity constraint on commodities into account. Note that due to the consideration of marketing strategies or practical conditions, quantities of commodities are often constrained in the real world. For example, the scarcity fallacy usually makes commodities more desirable [21], as a company may want to release limited edition commodities to boost customer desire. Another example is the case of promoting a concert, where the quantity of the concert tickets is constrained by the number of available seats. In this case, due to the schedules of the musicians or singers, it is hard to add more shows and the total number of concert tickets. Hence, selecting seeds to receive free commodities will reduce the quantities of commodities that can be sold. This makes the prior works unapplicable to determine the proper pricing and obtain maximum revenue. For example, suppose that the network shown in Figure 1 is considered, where the diffusion model with the monetary concept in the work [17] is adopted and the function  $F(x) = x$  is used to transform the effects of the social influence into the valuation concept. Without consideration of the commodity quantity, the maximum revenue is \$28 obtained at the pricing of \$7 with the seed group  $\{d, f\}$ . With the influence from  $\{d, f\}$ ,  $a$  and  $e$  are the first to purchase the commodity, since both their valuations  $\$2 + F(5) = \$7$  are equal to the pricing of \$7. Then,  $b$  and  $c$  will subsequently purchase the commodity due to their valuations  $\$0 + F(2 + 4 + 4) = \$10$  and  $\$3 + F(3 + 2 + 1) = \$9$ , respectively. However, if the quantity of commodities is limited to 4, the revenue will decrease to \$14 since two free commodities given to the seeds  $d$  and  $f$  should be taken off first. In this case, letting the commodity be priced at \$6 and choosing only  $d$  as the seed will be the best. Under the influence of  $d$ ,  $a$  first purchases the commodity since its valuation  $\$2 + F(5) = \$7$  is larger than the pricing of \$6. Then, both  $b$  and  $c$  purchase the commodities due to their valuations  $\$0 + F(2 + 4) = \$6$  and  $\$3 + F(3) = \$6$ , respectively. Finally, since  $a$ ,  $b$ , and  $c$  all pay for the commodity sold at the price of \$6, the maximum revenue is \$18. Therefore, when the quantity of commodities is limited, we argue the need of new approaches for the revenue maximization problem.

Specifically, we address the revenue maximization problem with a quantity constraint on commodities. Given a limited

quantity of commodities and a social network with monetary concept regarding to the commodity, the problem is to determine the pricing of the commodity and identify a small group of people, i.e., a seed group, to be the initial customers receiving freebies to help promote the commodity at the beginning, so that the total revenue is maximized. In this paper, we investigate this problem on the monetary-concept incorporated LT-model [17], and propose Algorithm PRUB (Pricing searching strategy using Revenue Upper Bound) to derive the optimal solutions. Then, we further revise PRUB to propose a heuristic Algorithm PRUB+IF (Pricing searching strategy using Revenue Upper Bound with Importance Feedback) for better efficiency. Experiments on real social networks show the good effectiveness of PRUB and PRUB+IF, demonstrating the revenue maximization with and without a quantity constraint differs from each other.

In summary, the contributions of this paper are:

1. We are the first to address the general case of the revenue maximization problem. Note that the previous work of the revenue maximization without a quantity constraint is a special case of our problem with the quantity set as the total number of individuals on the social network.
2. For the addressed problem, we proposed the optimal PRUB as well as a heuristic PRUB+IF.
3. Experiments on real social networks demonstrate the effectiveness of the proposed approaches for both general and special cases of the revenue maximization problem. Note even in the case without a commodity constraint, the proposed approach outperforms the state of the art approach that incorporates the monetary concept [17].

## 2. PROBLEM FORMULATION

To formulate the revenue maximization problem, in this section, we first describe the social network with the monetary concept, then explain how the social influences propagate over the network using the monetary-concept incorporated LT-model [17], and finally define the revenue and the revenue maximization problem formally.

Given a commodity, a social network with the monetary concept, referred to as a **monetizing social network**, is a weighted digraph  $G = (V, X, E, W, F)$ , where  $V$  is the set of individuals and for each individual  $v \in V$ , his/her inherent valuation  $\chi_v$  is carried in  $X$ . Usually, the valuation information  $X$  can be estimated and learned from questionnaires or historical sales data, with regard to the commodity [13]. For any individuals  $u, v \in V$ , if  $u$ 's purchase will directly encourage  $v$ 's desire for the commodity, an edge from  $u$  to  $v$ , denoted as  $e_{uv} \in E$  ( $E$  standing for the edge set), represents the existence of the influence, and  $w_{uv} \in W$  represents the influence strength as the weight on  $e_{uv}$ . Adopting the Concave Graph Model in the work [17], we consider a non-negative, non-decreasing, and concave function  $F: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  with  $F(0) = 0$  to transform the weights into the valuation concept. That is, given a set of individuals  $S$  directly exerting influences on an individual  $v$ ,  $v$ 's valuation towards a commodity will become  $\chi_v + F(\sum_{i \in S} w_{iv})$ .

Based on the monetary-concept incorporated LT-model [17], the propagation of social influences over the monetizing social network happens whenever individuals *adopt* the commodity. An individual *adopts* the commodity if and only

if 1) the pricing of the commodity is less than or equal to his/her valuation, or 2) this individual is an initial customer receiving a freebie. After an individual adopts the commodity, he/she will then exert influences on his/her out-neighbors to encourage them and raise their valuations for adoption. The following shows an example of social influences propagating over the monetizing social network.

*Example 1.* Given the monetizing social network in Figure 1 and the concave influence function  $F(x) = x$ , consider the promotion of a concert. Suppose that the pricing of the concert tickets is \$7 and the seed group is  $A = \{d\}$ . Influenced by  $d$ , the valuations of  $a$ ,  $b$ , and  $f$  are increased by  $F(5) = \$5$ ,  $F(4) = \$4$ , and  $F(2) = \$2$ , respectively. The valuations of  $a$ ,  $b$  and  $f$  thus become  $\$2 + \$5 = \$7$ ,  $\$0 + \$4 = \$4$ , and  $\$0 + \$2 = \$2$ , respectively. Then, note that as the valuation of  $a$  is equal to the pricing of the concert tickets \$7,  $a$  also adopts the concert ticket and exerts influences on its out-neighbors  $b$  and  $c$ . The valuations of  $b$  and  $c$  thus increase to  $\$0 + F(4 + 2) = \$6$  and  $\$3 + F(3) = \$6$ , respectively. After that, as no other individuals adopt the concert ticket, the influence propagation stops. Consequently, under  $d$ 's influence,  $a$  will also adopt the concert ticket, and the valuations of  $b$ ,  $c$  and  $f$  become \$6, \$6, and \$2, respectively.

Consider another example of the seed group  $A = \{d, f\}$ . Under the influence of  $d$  and  $f$ , the valuations of  $a$ ,  $b$ ,  $c$  and  $e$  are raised to \$7, \$4, \$4, and \$7, respectively. (Here  $d$ 's influence on  $f$  is neglected since  $f$  as an initial customer has already adopted the concert ticket.) Hence,  $a$  and  $e$  adopt the concert tickets and exert influences. Consequently, the valuations of  $b$  and  $c$  become:

$$\begin{aligned} b &: \$0 + F(4(\text{from } d) + 2(\text{from } a) + 4(\text{from } e)) = \$10 \\ c &: \$3 + F(1(\text{from } f) + 3(\text{from } a) + 2(\text{from } e)) = \$9 \end{aligned}$$

In the end,  $a$ ,  $b$ ,  $c$ , and  $e$  all adopt the concert tickets since their final valuations are larger than or equal to the pricing of the concert tickets \$7. ■

Here we define that the revenue comes from the number of people paying for the commodity and the pricing of the commodity. Recall Example 1 where the pricing is \$7 and the seed group is  $\{d, f\}$ , and suppose that the quantity of the concert tickets is 4. The revenue is  $\$7 \times 2 = \$14$ , since two concert tickets are used as free tickets given to  $d$  and  $f$ , and only two concert tickets are left for sale (even though  $a$ ,  $b$ ,  $c$ , and  $e$  all want to purchase these concert tickets).

Formally, given the quantity of commodities  $n$ , the pricing  $p$  of the commodity, and the seed group  $A$ , the revenue is defined as follows.

**Definition 1. Revenue function.** Given the quantity  $n$ , the revenue at the pricing of  $p$  with the seed group  $A$  is

$$R(n, p, A) = p \times \min\{|\sigma(A) \setminus A|, n - |A|\},$$

where  $\sigma(A) \supseteq A$  is the set of individuals adopting the commodity under the influences of  $A$ .

To increase the revenue, finding the proper pricing  $p$  and identifying a proper seed group  $A$  is important. Consider the same setting mentioned above, but with an empty seed group and the pricing set as \$1. The revenue is  $R(4, \$1, \emptyset) = \$1 \times 4 = \$4$ , which is less than  $R(4, \$7, \{d, f\}) = \$14$ . Since any company will expect to earn as higher revenue as possible given a fixed amount of commodities, in this paper, we are interested in looking for such pricing and a seed group

as initial customers that can bring maximum revenue. The proposed problem is formally defined as follows.

**Definition 2. The  $\text{RM}_{w/QC}$  Problem: Revenue Maximization with a Quantity Constraint.** Given a monetizing social network  $G = (V, X, E, W, F)$ , a set of input prices  $P \subseteq \mathbb{R}^+$ , and a quantity of commodities  $n$ , the problem is to determine the pricing  $p_{\max} \in P$  of the commodity and find a seed group  $A_{\max} \subseteq V$  as initial customers, where  $|A_{\max}| \leq n$ , such that the revenue  $R(n, p, A)$  is maximized, i.e.,  $(p_{\max}, A_{\max}) = \operatorname{argmax}_{p, A} R(n, p, A)$ .

Note that the revenue function  $R(n, p, A)$  is not single-peaked with respect to the pricing, which means the approaches such as the binary search and the gradient decent search are not applicable to finding out the optimal pricing. For example, following the setting of Example 1 and letting the quantity of the concert tickets be 4, we obtain the highest revenue at pricing of \$6, \$7, and \$8 as \$18, \$14, and \$16, respectively (with the seed groups  $\{d\}$ ,  $\{d, f\}$ , and  $\{d, e\}$ , accordingly). Obviously, with respect to the pricing,  $R(n, p, A)$  is not single-peaked.

Furthermore, the  $\text{RM}_{w/QC}$  problem can be proved to be NP-hard. To show this, we first introduce a special case of  $\text{RM}_{w/QC}$ , and then prove that this special case is NP-hard. As the special case is NP-hard,  $\text{RM}_{w/QC}$  is thus NP-hard.

**Definition 3. SpecialRM.** This problem is a special case of  $\text{RM}_{w/QC}$ , and asks for only one input price and a sufficient quantity of commodities for the population. That is, given  $G$ ,  $P = \{p\}$ , and  $n = |V|$ , SpecialRM is to determine a seed group  $A$  such that  $R(|V|, p, A)$  is maximized.

**THEOREM 1.** *The problem of SpecialRM is NP-hard.*

**PROOF.** We reduce the Minimum Vertex Cover (MVC) problem to the SpecialRM problem. Given every instance of MVC involving an undirected graph  $G' = (V', E')$ , we can construct a corresponding instance  $G = (V, X, E, W, F)$  of SpecialRM as follows. (1) Set  $V = V'$ ; (2) set  $X = \{\chi_v | \forall v \in V, \chi_v = 0\}$ ; (3) direct all edges in both directions, i.e., for each  $e_{uv} \in E'$ ,  $e_{uv} \in E$  and  $e_{vu} \in E$ ; (4) for each  $e_{uv} \in E$ , define  $w_{uv} = \frac{p}{d_{\text{in}}(v)}$ , where  $d_{\text{in}}(v)$  is the in-degree of  $v$  in  $G$ ; and (5) set  $F(x) = x$ . For the instance of SpecialRM, if there exists a minimum set  $A$  that maximizes the revenue, we can obtain another set  $A' = A \cup (V \setminus \sigma(A))$ . Then, it can be shown that  $A'$  is the minimum vertex cover of  $G'$  since

$$\sigma(A') \supseteq \sigma(A) \cup \sigma(V \setminus \sigma(A)) \supseteq \sigma(A) \cup (V \setminus \sigma(A)) = V.$$

Theorem 1 is thus proved. ■

### 3. ALGORITHM

In this section, for the revenue maximization problem, we first propose Algorithm PRUB (Pricing searching strategy using Revenue Upper Bound) that is able to derive the optimal solutions. Note that finding the optimal pair of  $p_{\max}$  and  $A_{\max}$  in the revenue maximization problem is exhausted. Then, for better efficiency, a heuristic Algorithm PRUB+IF (PRUB with Importance Feedback) is introduced following the framework of PRUB.

**Table 1: (a) The maximum valuations; (b) the upper bounds of maximum revenue.**

(a)		(b)	
$v$	$X_{\max}(v)$	$p$	$R_{\text{bound}}(4, p)$
$a$	\$8	\$1	\$4
$b$	\$10	\$2	\$8
$c$	\$9	\$3	\$12
$d$	\$4	\$4	\$16
$e$	\$7	\$5	\$20
$f$	\$4	\$6	\$24
		\$7	\$28
		\$8	\$24
		\$9	\$18
		\$10	\$10

### 3.1 Algorithm PRUB

To tackle the  $\text{RM}_{\text{w/QC}}$  problem, Algorithm PRUB is designed to obtain the optimal solutions. As mentioned previously, finding a proper pair of pricing and seed group is critical in this problem, since the revenue comes from the pricing of the commodity and the number of people paying for the commodity. A naïve approach for finding the optimal pair of pricing  $p_{\max}$  and seed group  $A_{\max}$  is to search at all  $p \in P$ , enumerate all seed groups  $A$ , and calculate the corresponding revenue  $R(n, p, A)$ . The rationale of Algorithm PRUB is based on this naïve approach and prunes the search space by 1) progressively filtering out non-candidate pricing and 2) deriving an upper bound of the size of seed groups for each price. We illustrate the two ideas of pruning in the following.

**Non-candidate Pricing Filtering.** The first idea is to prune the searching on non-candidate pricing. Motivated by the concern of only maximum revenue, this pruning derives an upper bound of maximum revenue at each price, and utilizes an achievable global revenue  $r_{\text{global}}$  that records the maximum revenue discovered so far to progressively filter out the pricing that will not result in higher revenue. That is, if an upper bound of maximum revenue at each price can be derived, PRUB then does not need to seek the seed group for maximum revenue at a specific price when the upper bound at this price is less than or equal to  $r_{\text{global}}$ . When  $r_{\text{global}}$  is updated and increased progressively after the successful searching of the seed group at each price, the non-candidate pricing is also progressively detected and filtered out.

In order to infer the upper bound of maximum revenue at a specific price, the maximum number of individuals who have potential for adopting the commodity at this price is needed. To find out whether or not an individual has this potential, it needs to estimate the valuation of the individual. Therefore, in the following, we first introduce the definitions of *maximum valuations* and *potential buyers* for defining the upper bound of maximum revenue.

**Definition 4. Maximum valuation.** The maximum valuation of an individual  $v$  is the valuation under the influences from all  $v$ 's in-neighbors, which is denoted as

$$X_{\max}(v) = \chi_v + F\left(\sum_{u \in v\text{'s in-neighbors}} w_{uv}\right).$$

**Definition 5. Potential buyer.** An individual  $v$  is regarded as a potential buyer at a specific price  $p$  if  $v$  has potential for adopting the commodity at  $p$ , i.e.,  $X_{\max}(v) \geq p$ .

**Definition 6. Upper bound of maximum revenue.** Given a quantity constraint  $n$ , the upper bound of maximum revenue at a price  $p$  is

$$R_{\text{bound}}(n, p) = p \times \min\{n, m_p\}, \quad (1)$$

where  $m_p$  is the number of potential buyers at  $p$ , i.e.,  $m_p = |\{v | v \in V, X_{\max}(v) \geq p\}|$ .

*Example 2.* Given the monetizing social network in Figure 1, the concave influence function  $F(x) = x$ , and a set of input prices  $P = \{p | p \in \mathbb{I}^+, 1 \leq p \leq 10\}$ , consider the promotion of a concert. Table 1(a) shows the maximum valuation of each individual. Suppose the quantity of the concert tickets is set as 4. The corresponding upper bounds of maximum revenue for all prices are listed in Table 1(b). ■

In addition, to have better effectiveness of pruning by utilizing the upper bound of maximum revenue, PRUB searches the prices in a descending order of the upper bounds. Then, once PRUB discovers a price, at which the upper bound is less than or equal to the achievable global revenue  $r_{\text{global}}$ , the searching for the following prices (including the current price) can be ignored. Therefore, in the above example, PRUB will start the examination of the prices from  $p = \$7$ ,  $p = \$6$ ,  $p = \$8$ ,  $\dots$ , and so on.

**Bound of Seed Group's Size.** The second idea for pruning search space is to avoid enumerating useless seed groups under a specific price. The inspiration is, under the consideration of a specific price  $p$ , it involves more than  $\frac{r_{\text{global}}}{p}$  quantities of commodities to be sold for finding the revenue higher than the achievable global revenue  $r_{\text{global}}$  obtained so far. When considering the price  $p$ , Algorithm PRUB then bounds the size of the seed groups as

$$|A| < n - \frac{r_{\text{global}}}{p}, \quad (2)$$

since the quantity of commodities is limited. Follow Example 2, and suppose the achievable global revenue  $r_{\text{global}} = \$10$ . In order to find the revenue higher than \$10, when searching at the price  $p = \$7$ , PRUB expects more than  $\frac{10}{7}$  concert tickets left for sale and enumerates only the seed groups of sizes less than or equal to 2 (since the quantity of the concert tickets is set as 4). The searching on the seed groups of sizes 3 and 4 is thus pruned.

Note that incorporating the two pruning methods is satisfactory to the accuracy. Therefore, Algorithm PRUB is still able to derive the optimal solutions, even though the search space is reduced. The details of PRUB are presented in Algorithm 1. In the following, we show an example to illustrate how PRUB figures out the optimal pair of  $p_{\max}$  and  $A_{\max}$ .

*Example 3.* Follow Example 2, and initialize  $p_{\max}$ ,  $A_{\max}$ , and  $r_{\text{global}}$  as \$0, an empty set, and \$0, respectively. According to the upper bounds of maximum revenue in Table 1(b), PRUB will consider the prices in the following order:  $p = \$7$ ,  $p = \$6$ ,  $p = \$8$ ,  $\dots$ , and so on.

Beginning from the price \$7, PRUB checks whether or not  $R_{\text{bound}}(4, \$7) > r_{\text{global}}$ . Since  $\$28 > \$0$ , PRUB looks for the maximum revenue at the price \$7 by enumerating all the seed groups whose size is bounded by Equation (2) (including the size 0). First, for the empty seed group, no individual adopts the concert ticket at \$7. So PRUB goes to the size 1. Then, PRUB lists all the seed groups of size 1 since

---

**Algorithm 1: PRUB**


---

**Input:** A monetizing social network  $G = (V, X, E, W, F)$ ; a set of input prices  $P$ ; a quantity of commodities  $n$ .

**Output:** The pricing  $p_{\max}$ ; the seed group  $A_{\max}$ .

```

1  $p_{\max} \leftarrow 0, A_{\max} \leftarrow \emptyset, r_{\text{global}} \leftarrow 0$ 
2 Derive  $R_{\text{bound}}(n, p)$  for all  $p \in P$ 
3 Sort all  $p \in P$  descendingly by  $R_{\text{bound}}(n, p)$ 
4 for  $p \in P$  do
5   if  $p$  is non-candidate pricing then
6     return  $p_{\max}, A_{\max}$ 
7   Enumerate all the seed groups whose size is bounded by
   Equation (2) (including the size 0)
8   Compute  $R(n, p, A)$  for those enumerated seed groups
9   if any  $R(n, p, A) > r_{\text{global}}$  then
10     $p_{\max} = p, A_{\max} = A, r_{\text{global}} = R(n, p, A)$ 
11 return  $p_{\max}, A_{\max}$ 

```

---

$1 < 4 - \frac{0}{7} = 4$ . Among all the seed groups of size 1, PRUB finds the highest revenue  $R(4, \$7, \{d\}) = \$7$  (referred to Example 1). As  $\$7$  is higher than  $r_{\text{global}} = \$0$ , PRUB updates  $p_{\max} = \$7$ ,  $A_{\max} = \{d\}$ , and  $r_{\text{global}} = \$7$ . Before seeking the seed groups of size 2 for maximum revenue, PRUB ensures  $2 < 4 - \frac{7}{7} = 3$ . After all seed groups of size 2 are enumerated, the highest revenue  $R(4, \$7, \{d, f\}) = \$14$  is found (referred to Example 1). Because  $\$14 > r_{\text{global}} = \$7$ , PRUB updates  $p_{\max} = \$7$ ,  $A_{\max} = \{d, f\}$ , and  $r_{\text{global}} = \$14$ . Later, note that  $3 > 4 - \frac{14}{7} = 2$ . The searching for the maximum revenue at the price  $\$7$  stops. The next price considered is  $\$6$ . A similar process is performed until PRUB finds such a price  $p$  that  $r_{\text{global}} \geq R_{\text{bound}}(4, p)$ . In the end, the maximum revenue is  $R(4, \$6, \{d\}) = \$18$ , where the pair of  $p_{\max} = \$6$  and  $A_{\max} = \{d\}$  is the optimal solution. ■

**THEOREM 2.** *The time complexity of PRUB is  $O(2^{|V|}|V|^2|P|)$ .*

**PROOF.** We first show that PRUB costs  $O(2^{|V|}|V|^2)$  time at a specific price for searching the maximum revenue. Given the quantity of commodities  $n$ , the largest size of the seed group is  $n$ . In total, there are thus  $\sum_{i=0}^n \binom{|V|}{i}$  seed groups. For a seed group of size  $i$ , there are  $i$  seeds exerting influences on the other  $O(|V| - i)$  individuals that are not seeds. Due to the influence cascade, it costs  $O((|V| - i)^2)$  time to figure out the influence spread of a seed group of size  $i$ . Therefore, the total time complexity for searching the maximum revenue by enumerating seed groups is

$$\begin{aligned}
& O\left(\sum_{i=0}^n \binom{|V|}{i} (|V| - i)^2\right) \\
&= O\left(\sum_{i=0}^n \frac{|V|!}{(|V| - i)!i!} (|V| - i)^2\right) \\
&= O\left(\sum_{i=0}^n \frac{|V|^2(|V| - 1)!}{(|V| - i - 1)!i!} - \frac{|V|(|V| - 1)(|V| - 2)!}{(|V| - i - 1)!(i - 1)!}\right) \\
&= O(|V|^2 \sum_{i=0}^n \binom{|V| - 1}{i} - |V|^2 \sum_{i=0}^n \binom{|V| - 2}{i - 1} + |V| \sum_{i=0}^n \binom{|V| - 2}{i - 1}) \\
&= O(2^{|V| - 1}|V|^2 - 2^{|V| - 2}|V|^2 + 2^{|V| - 2}|V|) \\
&= O(2^{|V|}|V|^2).
\end{aligned}$$

As searching at a price costs  $O(2^{|V|}|V|^2)$ , the total time complexity of PRUB is thus  $O(2^{|V|}|V|^2|P|)$ . Theorem 2 is proved. □

Note that the high complexity of Algorithm PRUB indicates the poor scalability of the optimal algorithm for real social networks which are usually large. Therefore, in the

following subsection, we propose a heuristic based on the framework of PRUB to obtain the feasible solutions on larger instances.

### 3.2 Algorithm PRUB+IF

In this section, the heuristic algorithm PRUB+IF (PRUB with Importance Feedback) is proposed as a feasible solution. The heuristic PRUB+IF differs from PRUB at the way of finding the most proper seed group at each price. Following the framework of PRUB, PRUB+IF introduces the concept of *pricing-sensitive importance*, accompanying the contribution feedback from the out-neighbors, for selecting seeds in a heuristic manner, instead of listing all seed groups.

The main idea of PRUB+IF is that, an individual with greater potential for making others adopt the commodity should be regarded as more important. Selecting the most important individuals as seeds greedily is able to lead to a feasible solution. Then, the question is how to evaluate one's potential for making others adopt the commodity. The intuition is about how many others will be encouraged and how much their valuations can be increased for approaching the pricing, under one's influence. We then accumulate these effects to calculate one's importance. Here only the effects on potential buyers should be included in the accumulation since only the potential buyers have possibilities of paying for the commodity. In addition, note that the individuals who newly adopt the commodity under one's (direct and indirect) influence will further spread their influences to encourage more others in adoption. In order to estimate one's importance more carefully, the effects through influence cascades should also be included in the accumulation. Therefore, PRUB+IF introduces the pricing-sensitive importance that sums up the consideration of the normalized weights, the influence propagation, and the potential buyers in the measurement of one's advantage in the commodity promotion at the given pricing. Then, the strategy is to select seeds in accordance with the pricing-sensitive importance.

In the following, we introduce the three key points, Normalized Weight, Feedback of Influence Propagation, and Potential-Buyer Filtering, in measuring the pricing-sensitive importance. Briefly, for each individual  $u$ , 1) the Normalized Weight is used to evaluate  $u$ 's importance according to  $u$ 's direct effect on another individual  $v$ 's adoption; 2) the Feedback of Influence Propagation is incorporated to evaluate  $u$ 's importance towards  $v$  by considering also the indirect effects through influence cascades; 3) the Potential-Buyer Filtering is considered to derive  $u$ 's pricing-sensitive importance from accumulating all  $u$ 's direct and indirect effects on all the other potential buyers.

#### 1. Normalized Weight.

By intuition, for commodity promotion, an individual  $u$ 's importance towards another individual  $v$  can be evaluated from how much  $v$ 's valuation approaches the pricing due to  $u$ 's effect. Inspired by this intuition, PRUB+IF calculates the normalized weight  $\hat{w}_{uv}$ , indicating  $u$ 's importance towards  $v$ 's adoption, as follows.

$$\hat{w}_{uv} =$$

$$\begin{cases} 0, & \text{if } p \leq X_A(v) \\ \min\{1, \frac{F(w_{uv} + \sum_{i \in \sigma(A)} w_{iv}) - F(\sum_{i \in \sigma(A)} w_{iv})}{p - X_A(v)}\}, & \text{otherwise} \end{cases}, \quad (3)$$

where  $X_A(v) = \chi_v + F(\sum_{i \in \sigma(A)} w_{iv})$  is  $v$ 's valuation of the commodity under the current seed group  $A$ 's effects. In the case of  $p \leq X_A(v)$ , it means  $v$  has adopted the commodity. So the normalized weight  $\hat{w}_{uv} = 0$ , indicating that  $u$ 's importance towards  $v$  is none due to no impact from  $u$  on  $v$ 's adoption of the commodity. For the other case, the normalized weight from  $u$  to  $v$  is bounded by 1, and is the portion of  $u$ 's impact on  $v$ 's adoption of the commodity.

*Example 4.* Given the monetizing social network in Figure 1 and the concave influence function  $F(x) = x$ , consider the promotion of a concert. Given  $p = \$7$  and  $A = \{\}$ , no one has adopted the concert tickets. Under this condition,  $X_{\{\}}(a) = \$2$  is less than the pricing of  $\$7$ , and the normalized weight from  $d$  to  $a$  is  $\hat{w}_{da} = \min\{1, \frac{\$5-\$0}{\$7-\$2}\} = 1$ . This means that selecting  $d$  into the seed group  $\{\}$  can make  $a$  adopt the concert ticket immediately. Similarly,

$$\hat{w}_{db} = \frac{4}{7}, \quad \hat{w}_{dc} = 0, \quad \hat{w}_{de} = 0, \quad \hat{w}_{df} = \frac{2}{7}. \quad \blacksquare$$

## 2. Feedback of Influence Propagation.

Note that the effect of influence propagation is important in the commodity promotion. To carefully estimate one's advantage in the promotion, PRUB+IF further takes the importance feedback from those individuals  $i$  into account on  $u$ 's importance, if  $i$  adopts the commodity due to  $u$ 's direct or indirect effect and thus devotes to the propagation of  $u$ 's effect. Therefore,  $u$ 's importance towards  $v$  will be derived recursively as follows.

$$IF^{(k+1)}(u, v) =$$

$$\begin{cases} \min\{1, IF^{(k)}(u, v) + \sum_{i \in V_u^{(k)}} \hat{w}_{iv}\}, & \text{if } u \neq v \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where  $IF^{(0)}(u, v) = \hat{w}_{uv}$  and  $V_u^{(k)}$  contains the individuals who newly adopt the commodity by  $u$ 's effects at  $k^{\text{th}}$  propagation, i.e.,  $V_u^{(k)} = \{i | i \in V, IF^{(k)}(u, i) = 1 \wedge \forall k' < k, IF^{(k')}(u, i) < 1\}$ . As the influence propagates, there are multiple updates for  $u$ 's importance towards  $v$ . When the propagation stops, i.e.,  $V_u^{(k)} = \emptyset$ , the final importance of  $u$  towards  $v$  is denoted as  $IF(u, v)$ . In other words, Equation (4) evaluates the advantage of selecting  $u$  into the current seed group from the aspect of  $v$ 's adoption, by considering the direct effect  $IF^{(0)}(u, v) = \hat{w}_{uv}$  derived by Equation (3) and the indirect effects  $\sum_k \sum_{i \in V_u^{(k)}} \hat{w}_{iv}$  propagated

through individuals  $i$ . In addition, Equation (4) ensures that the importance of  $u$  towards  $v$  is bounded by 1, even though there may be many direct and indirect impacts from  $u$  to  $v$ .

*Example 5.* Follow Example 4 and consider the individual  $d$ . Initially,  $V_d^{(0)} = \{a\}$ , because  $IF^{(0)}(d, a) = \hat{w}_{da} = 1$ .  $IF^{(1)}(d, \cdot)$  for all individuals (excluding  $d$  itself) are thus calculated as follows.

---

## Algorithm 2: PRUB+IF

---

**Input:** A monetizing social network  $G = (V, X, E, W, F)$ ; a set of input prices  $P$ ; a quantity of commodities  $n$ .

**Output:** The pricing  $p_{\max}$ ; the seed group  $A_{\max}$ .

```

1  $p_{\max} \leftarrow 0, A_{\max} \leftarrow \emptyset, r_{\text{global}} \leftarrow 0$ 
2 Derive  $R_{\text{bound}}(n, p)$  for all  $p \in P$ 
3 Sort all  $p \in P$  descendingly by  $R_{\text{bound}}(n, p)$ 
4 for  $p \in P$  do
5   if  $p$  is non-candidate pricing then
6     return  $p_{\max}, A_{\max}$ 
7    $A \leftarrow \emptyset$ 
8   Compute  $R(n, p, A)$ 
9   if  $R(n, p, A) > r_{\text{global}}$  then
10      $p_{\max} = p, A_{\max} = A, r_{\text{global}} = R(n, p, A)$ 
11   while  $|A| < n - \frac{r_{\text{global}}}{p}$  do
12     for  $u \in V$  has not adopted the commodity do
13       Compute  $\Psi(u)$  according to Equation (5)
14        $s \leftarrow$  the individual with the greatest  $\Psi(\cdot)$ 
15        $A \leftarrow A \cup \{s\}$ 
16       Compute  $R(n, p, A)$ 
17       if  $r > r_{\text{global}}$  then
18          $p_{\max} = p, A_{\max} = A, r_{\text{global}} = R(n, p, A)$ 
19 return  $p_{\max}, A_{\max}$ 

```

---

$$\begin{aligned} IF^{(1)}(d, a) &= \min\{1, 1 + \hat{w}_{aa}\} = 1 \\ IF^{(1)}(d, b) &= \min\{1, \frac{4}{7} + \hat{w}_{ab}\} = \frac{6}{7} \\ IF^{(1)}(d, c) &= \min\{1, 0 + \hat{w}_{ac}\} = \frac{3}{4} \\ IF^{(1)}(d, e) &= \min\{1, 0 + \hat{w}_{ae}\} = 0 \\ IF^{(1)}(d, f) &= \min\{1, \frac{2}{7} + \hat{w}_{af}\} = \frac{2}{7} \end{aligned}$$

Since no other individuals newly adopt the concert ticket, i.e.,  $V_d^{(1)} = \emptyset$ , the propagation stops. Finally,  $IF(d, \cdot) = IF^{(1)}(d, \cdot)$  are as above.  $\blacksquare$

## 3. Potential-Buyer Filtering.

Once the advantage of selecting an individual as a seed can be estimated from the aspect of another's adoption, intuitively, the importance of an individual can be derived by summing up his/her importance towards all the others. However, counting one's importance towards all the others may be misleading, since the advantage estimated from those who will never adopt the commodity at this price is also counted. To avoid overestimating the importance of an individual in the promotion, PRUB+IF thus incorporates the concept of potential buyers into the pricing-sensitive importance which is formally defined below.

$$\Psi(u) = \sum_{i \in V_{\text{potential}}} IF(u, i) \quad (5)$$

where  $V_{\text{potential}}$  is the set of potential buyers at this price, i.e., the individuals whose maximum valuation is larger than or equal to the price (the same as introduced in Section 3.1). As a result, only if an individual is a potential buyer who has potential for adopting the commodity at the price, the increase of his/her valuation will be regarded as important.

*Example 6.* Follow Example 5. Since the potential buyers at the pricing of  $\$7$  are  $a, b, c$ , and  $e$  (referred to Table 1(a)), the pricing-sensitive importance of  $d$  is  $\Psi(d) = 1 + \frac{6}{7} + \frac{3}{4} + 0 = \frac{73}{28}$ .  $\blacksquare$

The algorithm of PRUB+IF is presented in Algorithm 2. PRUB+IF uses the framework of PRUB as a base, and differs from PRUB at the strategy for seed selection (Algorithm 2 Lines 12-15). In the following, we go through the details of PRUB+IF. Similarly to PRUB, PRUB+IF initializes  $p_{\max}$ ,  $A_{\max}$ , and  $r_{\text{global}}$ , compute the upper bounds of maximum revenue for all prices, and then considers the prices one-by-one in a descending order of the upper bounds of maximum revenue (Lines 1-3). At a specific price  $p$ , PRUB enumerates all seed groups while PRUB+IF iteratively picks up a seed until the size of the seed group  $A$  reaches the bound in Equation (2) (Line 11). Specifically, when considering a price  $p$ , PRUB+IF first checks whether or not  $p$  is non-candidate pricing that has no chance to yield higher revenue than the achievable global revenue  $r_{\text{global}}$  (Line 5). If  $p$  passes the pruning, PRUB+IF first calculates the revenue regarding to the empty seed group (Lines 7-8). After that, PRUB+IF expands the seed group by one seed at a time according to the pricing-sensitive importance  $\Psi(\cdot)$  in Equation (5) (Line 13). Each time the individual who has not yet adopted the commodity and has the greatest pricing-sensitive importance will be selected into the current seed group (Lines 12-15). In the case of the newly obtained revenue higher than  $r_{\text{global}}$ , PRUB+IF updates  $p_{\max}$ ,  $A_{\max}$ , and  $r_{\text{global}}$  similarly to PRUB (Lines 17-18). Once the size of the seed group reaches the bound in Equation (2) (Line 11), PRUB+IF considers the next price. After all prices are examined,  $p_{\max}$  and  $A_{\max}$  are returned as the solution. The following example demonstrates the searching process of PRUB+IF.

*Example 7.* Follow Example 2, and initialize  $p_{\max}$ ,  $A_{\max}$ , and  $r_{\text{global}}$  as \$0, an empty set, and \$0, respectively. According to the upper bounds of maximum revenue in Table 1(b), the first price considered is \$7. Since  $R_{\text{bound}}(4, \$7) = \$28 > r_{\text{global}} = \$0$ , PRUB+IF tries to seek for higher revenue at the price \$7. First, for  $A = \{\}$ , there is no revenue since the inherent valuations of all individuals are less than \$7. After that, for selecting the first seed, PRUB+IF computes the pricing-sensitive importance of all individuals in a similar way as shown in Example 6:

$$\begin{aligned} \Psi(a) &= \frac{29}{28}, & \Psi(b) &= \frac{1}{5}, & \Psi(c) &= 0, \\ \Psi(d) &= \frac{73}{28}, & \Psi(e) &= \frac{15}{14}, & \Psi(f) &= \frac{65}{28}. \end{aligned}$$

Hence, the first seed selected is  $d$  and the corresponding revenue will be  $R(4, \$7, \{d\}) = \$7$  (referred to Example 1). As  $\$7 > r_{\text{global}} = \$0$ , PRUB+IF performs the updates of  $p_{\max} = \$7$ ,  $A_{\max} = \{d\}$ , and  $r_{\text{global}} = \$7$ . Until now, the process for the first seed finishes, and PRUB+IF checks whether or not to continue selecting the second seed. Since  $|A| = |\{d\}| = 1 < n - \frac{r_{\text{global}}}{p} = 4 - \frac{7}{7} = 3$ , the pricing-sensitive importance of all individuals, except the individuals who have adopted the concert tickets, i.e.,  $a$  and  $d$ , is calculated with respect to the current seed group  $A = \{d\}$ :

$$\Psi(b) = 0, \quad \Psi(c) = 0, \quad \Psi(e) = 2, \quad \Psi(f) = 3.$$

Accordingly,  $f$  will be selected as the second seed and the revenue is  $R(4, \$7, \{d, f\}) = \$14$  (referred to Example 1), which leads to the updates of  $p_{\max} = \$7$ ,  $A_{\max} = \{d, f\}$ , and  $r_{\text{global}} = \$14$ . Until now, the process for the second seed finishes, and PRUB+IF checks whether or not to continue selecting the third seed. However,  $|A| = 2$  reaches the bound  $n - \frac{r_{\text{global}}}{p} = 4 - \frac{14}{7} = 2$  so that the searching at the price \$7 stops.

The searching at the remaining prices  $p = \$6$ ,  $p = \$8$ ,  $\dots$ , and so on, is performed in the same manner as that at the price \$7. In the end, PRUB+IF will report the pricing of \$6 and the seed group  $\{d\}$  as the answer. ■

**THEOREM 3.** *The time complexity of PRUB+IF is  $O(n|V|^3|P|)$ .*

**PROOF.** We first show that PRUB+IF costs  $O(n|V|^3)$  time at a specific price for searching the maximum revenue. Given the quantity of commodities  $n$ , the largest size of the seed group is  $n$ . When  $i$  seeds have been selected, it takes  $O((|V| - i)^2)$  time to compute the pricing-sensitive importance for each of the other  $O(|V| - i)$  nodes. Therefore, the total time complexity for searching the maximum revenue is

$$O\left(\sum_{i=1}^{n-1} (|V| - i)^3\right) = O(n|V|^3).$$

As searching at a price costs  $O(n|V|^3)$ , the total time complexity of PRUB+IF is thus  $O(n|V|^3|P|)$ . Theorem 3 is proved. □

## 4. EXPERIMENTS

In this section, we first provide the evidence that the revenue maximization problem with a quantity constraint differs from that without a quantity constraint. Besides, we also show the approximation performance of PRUB+IF to PRUB. In order to demonstrate the effectiveness of PRUB+IF, we further conduct experiments on larger real social networks and compare PRUB+IF with other heuristics. All the programs are implemented in Java with version 1.6.0\_27. The experiments are performed on a quad-core Intel® Xeon® X5450 3.00GHz PC with 8GB RAM using Ubuntu 8.04.2.

**Datasets.** We use three real social networks as follows.

1) The dataset *highschool* (sampled from the dataset [6] due to poor scalability of PRUB and the compared approach) is a friendship network for the boys in a small highschool in Illinois.<sup>1</sup> The nodes are boys in the highschool, and each directed edge represents a boy choosing another as a friend in questionnaires. 2) The dataset *digg* [4] is a reply network of Digg.<sup>2</sup> Each node is a user in Digg and each directed edge stands for a user replying to another. 3) The dataset *facebook* [22] is the communication network of Facebook.<sup>3</sup> The nodes are the users in Facebook while each directed edge represents that a user posts an article on another user's wall. For these datasets, the self-loops and isolated nodes are removed, and the edge weights are derived from the number of choices, replies, and articles of one to another, respectively. Both the degree distributions and weight distributions of all the datasets follow the well-known power law [5]. More details are shown in the left five columns of Table 2.

**Valuation Distributions.** Note that there is lack of valuation information in these publicly available datasets. A naïve way is to learn the valuation information through questionnaires or historical sales data toward a specific commodity. However, note that the valuation distributions vary with commodities. In this paper, we then explore the performance under different valuation distributions as follows.

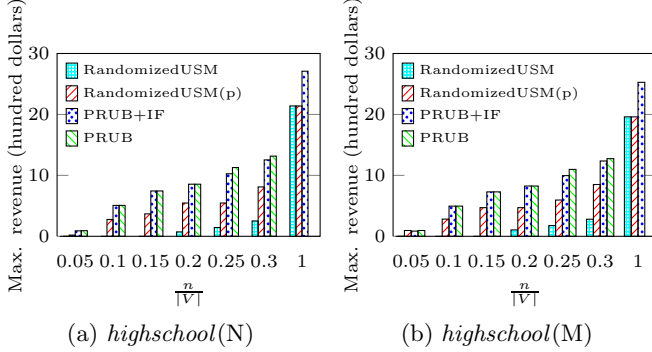
<sup>1</sup><http://moreno.ss.uci.edu/data.html#high>

<sup>2</sup><http://www.public.asu.edu/~mdechoud/datasets.html> (The reply system has been out of service since 2012, but the data were collected in 2008.)

<sup>3</sup><http://socialnetworks.mpi-sws.org/data-wosn2009.html>

**Table 2: Summaries of real datasets and the parameters of each valuation distribution.**

Dataset	#nodes	#edges	Avg. degree	Avg. weight	Avg. clustering coeff.	Normal		M-shape			
						$\mu$	$\sigma^2$	$\mu$	$\sigma^2$	$\mu$	$\sigma^2$
<i>highschool</i>	50	140	5.60	37.29	0.4754	10	8.16	4	1.78	16	1.78
<i>digg</i>	30,360	85,155	5.61	6.07	0.0053	5	2.04	2	0.44	8	0.44
<i>facebook</i>	45,813	183,412	8.01	4.66	0.1106	5	2.04	2	0.44	8	0.44


**Figure 2: The comparison for maximum revenue.**

- Normal distribution (N): Most people have similar valuations and a few people have extremely higher or lower valuations toward a commodity.
- M-shape distribution (M) [18]: People do not have consensus on the value of a commodity. One group of people have higher valuations and the other group of people have lower valuations toward the commodity.

In our experiments, we implement the M-shape distribution using two Normal distributions, where the valuation of each individual follows one of the Normal distributions with equal probabilities. The parameters for generating the valuations are listed in the right three columns of Table 2.

**Concave influence function.** As introduced in Section 2, any non-negative, non-decreasing, and concave function can be applied to the concave influence function. In our experiments, we consider the function  $F(x) = x$  to transform the weights into the valuation concept for simplicity.

**Input prices.** By considering different currencies, costs, and cultures in different countries, the pricing of a commodity may be different. For simplicity, we set the prices as input parameters in the experiments. For *highschool*, the input prices are all integers (as the unit of the dollars) in  $[1, 300]$ , due to the smaller size of the network. For *digg* and *facebook*, the input prices are all integers in  $[1, 2000]$ .

## 4.1 Consideration of Quantity Constraints

In this subsection, we first show that the consideration of a quantity constraint in the revenue maximization problem makes it differ from the original problem by comparing PRUB with the previous work [17], referred to as RandomizedUSM in this paper. Due to poor scalability of PRUB and RandomizedUSM, these experiments are conducted on the small dataset *highschool* with Normal and M-shape valuation distributions. In addition, note that since RandomizedUSM does not consider a quantity constraint, we derive the revenue with two kinds of post-processes as follows.

- RandomizedUSM: Given the solution pair of  $p$  and  $A$  RandomizedUSM reports, this post-process considers the quantity constraint  $n$  by calculating the quantities left for sale and computing the revenue as  $p \times$

$\min\{|\sigma(A) \setminus A|, n - |A|\}$ , where  $\sigma(A) \supseteq A$  is the set of individuals adopting the commodity under  $A$ 's effects.

- RandomizedUSM(p): This post-process considers the quantity constraint  $n$  at each price and then returns the highest revenue among all prices. For each price  $p$  and the corresponding seed group  $A_p$  obtained by RandomizedUSM, the revenue is  $p \times \min\{|\sigma(A_p) \setminus A_p|, n - |A_p|\}$ , where  $\sigma(A_p) \supseteq A_p$  is the set of individuals adopting the commodity under  $A_p$ 's effects.

Figure 2 shows the maximum revenue reported by RandomizedUSM, RandomizedUSM(p), PRUB+IF and PRUB with respect to different ratios of the quantity of commodities to the number of population, i.e.,  $\frac{n}{|V|}$ . Accordingly, the revenue reported by RandomizedUSM and RandomizedUSM(p) is, on average, less than 70% of that obtained by PRUB. Moreover, note that RandomizedUSM has the poorest performance and even brings zero revenue at the cases of the supply ratio below 0.2. These results demonstrate that the capability of the previous approach [17] (even with the post-processes) is limited for the RM<sub>w/QC</sub> problem. On the other hand, even for the special case that the quantity is not constrained, i.e., the ratio of the quantity of commodities to the number of population  $\frac{n}{|V|} = 1$ , PRUB+IF still outperforms RandomizedUSM and RandomizedUSM(p). This is because PRUB+IF selects seeds in order of the pricing-sensitive importance while RandomizedUSM and RandomizedUSM(p) consider the seeds in arbitrary order.

On the same datasets, we then also explore the approximation of PRUB+IF to PRUB (which derives the optimal solutions). According to the results in Figure 2, the performance of PRUB+IF is very close to that of PRUB, i.e., about 96% on average, and outperforms both RandomizedUSM and RandomizedUSM(p). This implies that the proposed pricing-sensitive importance of PRUB+IF is highly effective to identify the crucial individuals as seeds.

## 4.2 Performance of the PRUB+IF heuristic

In this subsection, we conduct experiments on larger datasets *digg* and *facebook* to further show the performance of Algorithm PRUB+IF in terms of effectiveness, pricing, and efficiency. Here PRUB+IF is compared with two other heuristics PRUB+R and PRUB+SW. These two heuristics also take the PRUB approach as the framework while applying different strategies for seed selection. Specifically, PRUB+R (R standing for Random) selects seeds randomly; PRUB+SW (SW meaning Sum of the Weights) always picks up the individual with the maximum sum of out-weights from those who have not yet adopted the commodity. Besides, in order to demonstrate the advantage of each part of PRUB+IF, the heuristics PRUB+N, PRUB+F, and PRUB+P represent the approaches that incorporate Normalized Weight, Feedback of Influence Propagation, and Potential-buyer Filtering, respectively, for seed selection.

**Effectiveness.** Here we use the maximum revenue obtained without consideration of social influences (NoSocial), i.e., the adoption of an individual only depends on his/her in-



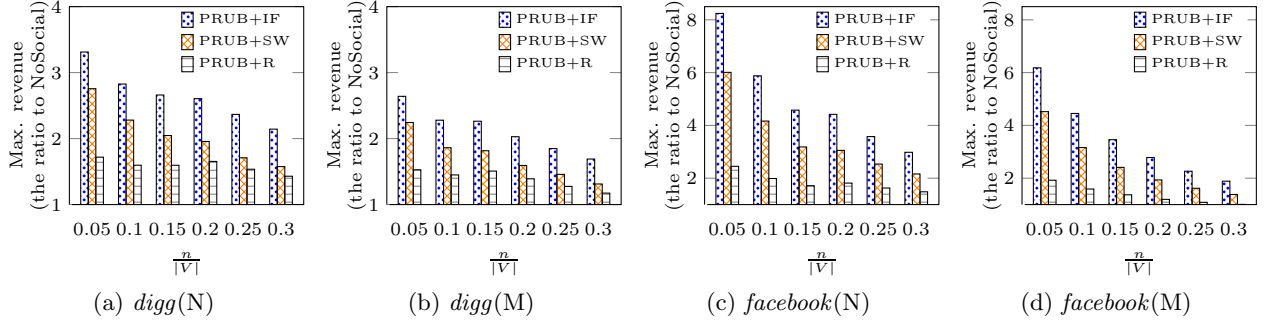


Figure 3: The effectiveness comparison.

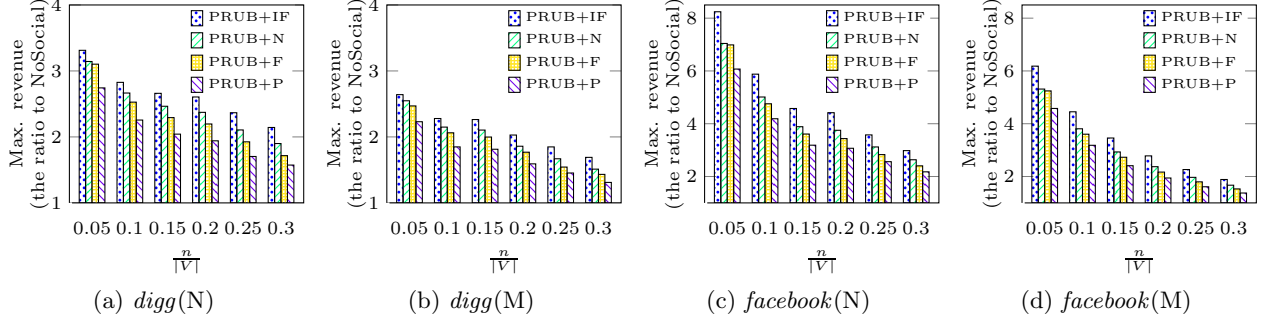


Figure 4: The advantage of each part of PRUB+IF.

herent valuation, as the baseline, and show the effectiveness of the heuristics in ratios to the baseline. Figure 3 reports the effectiveness with respect to different ratios of the quantity of commodities to the number of population, i.e.,  $\frac{n}{|V|}$ . For these results, we have two observations as follows. First, on *digg* and *facebook* with both Normal and M-shape valuation distributions, PRUB+IF is the best, and PRUB+SW outperforms PRUB+R only in the cases of small  $\frac{n}{|V|}$ . The reason is that, to estimate the advantage of selecting an individual in the promotion, PRUB+IF introduces the pricing-sensitive importance, which considers the normalized effects of the influence propagated through the social network, while PRUB+SW only takes the sum of weights on one's direct out-neighbors. PRUB+R picks the seeds blindly. Second, as  $\frac{n}{|V|}$  decreases, the performance of PRUB+IF becomes even more noticeable (compared to PRUB+SW and PRUB+R). This is because the number of potential buyers reduces when the pricing of the commodity increases due to scarcity. The less the number of potential buyers is, the more challenging the seed selection is. Hence, the results demonstrate that the strategy of PRUB+IF for the seed selection outperforms those of PRUB+SW and PRUB+R.

In order to demonstrate the advantage of each part of PRUB+IF, Figure 4 shows the effectiveness of PRUB+IF, PRUB+N, PRUB+F, and PRUB+P with respect to different ratios of the quantity of commodities to the number of population, i.e.,  $\frac{n}{|V|}$ . Accordingly, the three parts of PRUB+IF are more or less all equally important.

**Pricing.** In addition to the demonstration of effectiveness by maximum revenue, whether or not the suggested pricing complies with the common sense is also an interesting question. For answering this question, we plotted the trend of the pricing suggested by PRUB+IF, PRUB+SW, PRUB+R, and NoSocial with respect to different ratios of the quantity of commodities to the number of population, i.e.,  $\frac{n}{|V|}$ ,

in Figure 5. Accordingly, PRUB+IF suggests higher pricing in most cases than the other approaches to gain higher maximum revenue (as shown in Figure 3) in all cases. This implies that PRUB+IF performs the most elegant strategy for seed selection. The smart utilization of word-of-mouth effects can help the promotion of commodities in higher pricing, which complies with the common sense. Furthermore, for all methods, there is an upward trend in the pricing as  $\frac{n}{|V|}$  decreases. The less the quantity is, the more precious the commodity could be, which also complies with the common sense. There are some exceptions to this pricing trend: PRUB+IF at 0.25 on *digg*(M), and both PRUB+SW and PRUB+R at 0.25 or 0.3 on *facebook*(N). The reason is that a quantity of supply is used as freebies. This may allow higher pricing in some cases by leaving less quantities for sale. The tiny example in Figure 6 explains such a case in more details. ( $F(x) = x$  is used as the concave influence function.) For the case of the quantity of the concert tickets as 2, the maximum revenue is  $R(2, \$3, \emptyset) = \$6$  (i.e.,  $b$  and  $c$  adopt the concert tickets). Now consider that the quantity of the concert tickets is increased to 3. The maximum revenue is  $R(3, \$7, \{a, c\}) = \$7$  (i.e.,  $b$  adopts the concert ticket under the impact of  $a$  and  $c$ ). The pricing is higher even though the quantity increases. This is because 2 out of 3 concert tickets are used as free tickets and only 1 concert ticket is left for sale. Hence, according to these results, PRUB+IF is capable of suggesting elegant marketing strategies.

**Efficiency.** Now we compare the efficiency of the three heuristics. Figure 7 shows the runtime with respect to different ratios of the quantity of commodities to the number of population, i.e.,  $\frac{n}{|V|}$ , on *digg*(N), *digg*(M), *facebook*(N), and *facebook*(M). PRUB+IF spends more execution time than PRUB+SW and PRUB+R in all cases, since PRUB+IF considers the influence propagation in the derivation of pricing-sensitive importance for seed selection. PRUB+R generally

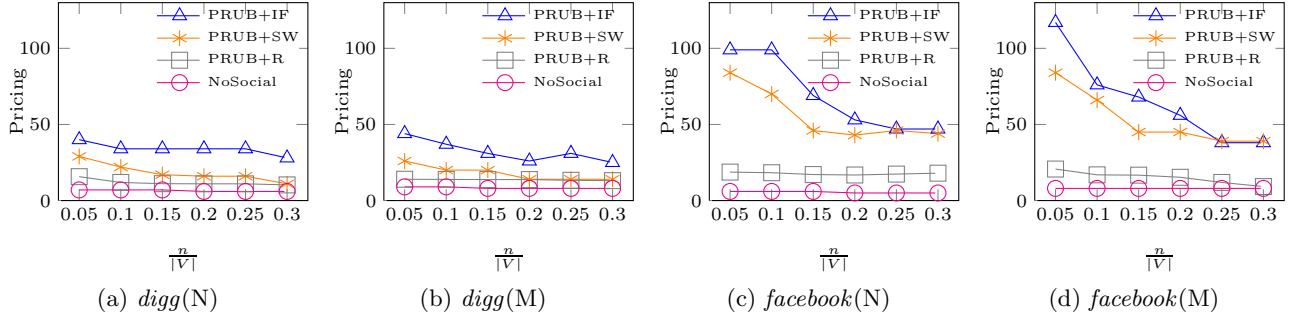


Figure 5: The pricing comparison.

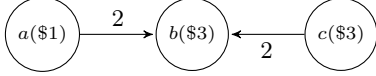


Figure 6: An example for the exception to the pricing trend.

spends a little bit more time than PRUB+SW due to the blind seed selection (that makes PRUB+R need to search more prices for finding maximum revenue). Nevertheless, the runtime of all the three heuristics is not sensitive to  $\frac{n}{|V|}$ . For PRUB+IF, other factors such as the local structure of potential buyers and the influence cascades will also affect the runtime.

## 5. RELATED WORKS

**Influence Maximization.** The influence maximization problem aims at identifying a group of seeds so that the number of the active people is the largest [3, 7, 15, 19]. To the best of our knowledge, the first study could be traced to the work [7], where Domingos et al. studied the influence maximization problem as an algorithmic problem. Then, Kempe et al. [15] formally modeled the problem as a discrete optimization, which is to identify  $k$  seeds for maximizing the influence spread over the social network. Meanwhile, Kempe et al. also showed that the influence maximization problem is NP-hard under both the Linear Threshold (LT) and the Independent Cascade (IC) models, and proposed a greedy strategy with approximation guarantees. Specifically, in the LT model, the influences on an individual are cumulative, and each individual has his/her own activation threshold. An individual becomes active if the sum of weights from his/her active neighbors is larger than or equal to his/her activation threshold. Then, this active individual will also propagate influences to his/her inactive neighbors. In the IC model, each influence an individual receives is independent. For each influence, there is one and only one chance to activate the individual with a probability. If the individual becomes active, he/she further spreads his/her influence in the same manner. In the work [3], Chen et al. showed that, given seeds, computing the influence spread under the LT model is #P-hard, and provided a scalable heuristic based on the fast computation for directed acyclic graphs (DAGs). Singer [19] concerned the inherent costs of making individuals as seeds and followed the incentive compatible mechanism to extract the true information about individuals' costs for identifying and rewarding the seeds. As mentioned in Section 1, all these influence maximization problems do not take the monetary effects on people's purchase decisions into account, and thus differ from our problem.

**Revenue Maximization.** The problem of revenue maximization is to derive the marketing strategy that brings the optimal revenue under the social influence, and was first addressed by Hartline et al. [11]. As a solution, Hartline et al. [11] proposed a marketing strategy, named influence-and-exploit (IE), to tackle the problem in two steps. In the influence step, IE identifies a group of customers as seeds. After that, in the exploit step, IE visits the remaining customers in a random order and offers each of them the optimal (myopic) pricing of the product in order to maximize the revenue. The decision whether or not an individual purchases the product depends on the pricing offered and the individual's valuation, which takes the social influences into consideration. Hartline et al. [11] also discussed the inferences of the valuations from social influences by Uniform Additive Model, Symmetric Model, and Concave Graph Model. Later, Fotakis et al. [9] proposed a polynomial-time approach to approximate the maximum revenue in the Uniform Additive Model. Another work [16] distinguishes between activation and adoption. When an individual is active by the social influence, he/she will then determine whether or not to adopt the product with respect to the pricing and his/her valuation. The marketing strategy of all these works is to determine customized pricing for different customers.

Mirroknii et al. [17] argued that offering the product at fixed pricing is more reasonable since offering customized pricing is hard to implement in the real world. Hence, their objective is to discover a seed group with the fixed pricing to maximize the revenue. Mirrokni et al. [17] adopted the Concave Graph Model proposed in the work [11] to incorporate the social influence into the valuation, and presented a  $(1/2)$ -approximation based on the randomized linear time approximation algorithm for the Unconstrained Submodular Maximization problem.

However, all the works above assumed that there is an unlimited supply of commodities, whereas we consider the scenario with limited supply such as limited edition commodities and concert tickets. As shown in Sections 1 and 4.1, it is unapplicable for these works to determine the proper marketing strategies.

## 6. CONCLUSION

In this work, we addressed the revenue maximization problem with a quantity constraint on a monetizing social network. To maximize the revenue, the marketing strategy is to determine the pricing of the commodity and find a seed group of individuals as the initial customers. Hence, we provided Algorithm PRUB to derive the pricing and the seed group for the optimal revenue. For better efficiency, we fur-

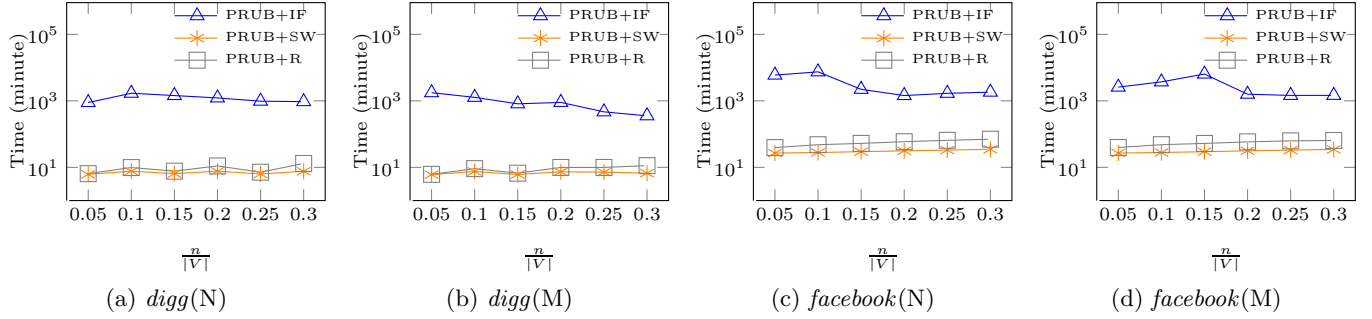


Figure 7: The runtime comparison.

ther proposed Algorithm PRUB+IF as a heuristic algorithm based on the framework of PRUB for a feasible solution. The experiments on real datasets with different valuation distributions showed the effectiveness of PRUB and PRUB+IF.

## 7. REFERENCES

- [1] D. Arthur, R. Motwani, A. Sharma, and Y. Xu. Pricing Strategies for Viral Marketing on Social Networks. In *WINE*, 2009.
- [2] S. E. Asch. Effects of Group Pressure on the Modification and Distortion of Judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 177-190). Pittsburgh, PA: Carnegie Press, 1951.
- [3] W. Chen, Y. Yuan, and L. Zhang. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. In *IEEE ICDM*, 2010.
- [4] M. D. Choudhury, H. Sundaram, A. John, D. D. Seligmann. Social Synchrony: Predicting Mimicry of User Actions in Online Social Media. In *IEEE CSE*, 2009.
- [5] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law Distributions in Empirical Data. In *SIAM review* 51.4, 2009.
- [6] J. S. Coleman. Introduction to Mathematical Sociology. *London Free Press Glencoe*, 1964.
- [7] P. Domingos and M. Richardson. Mining the Network Value of Customers. In *ACM SIGKDD*, 2001.
- [8] D. Easley and J. Kleinberg. Information Cascades. In *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (chap. 16, pp. 483-508). Cambridge University Press, 2010.
- [9] D. Fotakis and P. Siminelakis. On the Efficiency of Influence-and-Exploit Strategies for Revenue Maximization under Positive Externalities. In *WINE*, 2012.
- [10] H. Fromlet. Predictability of Financial Crises: Lessons from Sweden for Other Countries. *Business Economics* 47.4, 2012.
- [11] J. Hartline, V. S. Mirrokni, and M. Sundararajan. Optimal Marketing Strategies over Social Networks. In *ACM WWW*, 2008.
- [12] B. Hu, M. Jamali, and M. Ester. Learning the Strength of the Factors Influencing User Behavior in Online Social Networks. In *IEEE ASONAM*, 2012.
- [13] S. Y. Jung, J.-H. Hong, and T.-S. Kim. A Statistical Model for User Preference. In *IEEE TKDE*, 2005.
- [14] M. L. Katz and C. Shapiro. Network Externalities, Competition, and Compatibility. *American Economic Review* 75: 424-440, 1985.
- [15] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the Spread of Influence through a Social Network. In *ACM SIGKDD*, 2003.
- [16] W. Lu and L. V. Lakshmanan. Profit Maximization over Social Networks. In *IEEE ICDM*, 2012.
- [17] V. S. Mirrokni, S. Roch, and M. Sundararajan. On Fixed-Price Marketing for Goods with Positive Network Externalities. In *WINE*, 2012.
- [18] K. Ohmaa. The Impact of Rising Lower-Middle Class Population in Japan. Tokyo: Kodan-sha Publishing Company, 2006.
- [19] Y. Singer. How to Win Friends and Influence People, Truthfully: Influence Maximization Mechanisms for Social Networks. In *ACM WSDM*, 2012.
- [20] A. Sundararajan. Local Network Effects and Complex Network Structure. *The BE Journal of Theoretical Economics* 7(1), 2007.
- [21] S. Worchel, J. Lee, and A. Adewole. Effects of Supply and Demand on Ratings of Object Value. *Journal of Personality and Social Psychology* 32.5 (1975): 906.
- [22] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the Evolution of User Interaction in Facebook. In *ACM WOSN*, 2009.